ML PROJECT

Cancer Detection using Deep Learning

UI21EC35

Introduction

Cancer remains one of the leading causes of death worldwide, with early detection playing a pivotal role in improving patient outcomes and survival rates. Traditional methods of cancer detection often rely on invasive procedures and manual interpretation of medical imaging data, which can be time-consuming and prone to human error.

The machine learning project aims to predict whether a given photo of a tongue indicates the presence of cancer or not. The project involves training a machine learning model on a dataset comprising images of tongues from both cancer and non-cancer patients.

The model learns to identify patterns and features in the images that are indicative of cancerous or healthy conditions. Once trained, the model can analyse new tongue images and make predictions about the likelihood of cancer presence. This project holds promise for early detection and diagnosis of oral cancer, potentially improving patient outcomes through timely intervention.

Problem Statement: - The need for accurate and efficient methods for cancer detection using medical imaging. advanced techniques to enhance the accuracy, speed, and reliability of cancer detection.

Background Research

What is need of this Cancer-detection project??

In recent years, the global incidence of cancer has remained high. Tens of millions of people are newly diagnosed with various types of cancer every year. At the same time, millions to nearly tens of millions of people around the world are killed by various types of cancer.

In many cases, early detection is the key to improving the survival rate of cancer patients. For detection of the Cancer, Imaging techniques include B-ultrasound, X-ray, computed tomography (CT), magnetic resonance imaging (MRI), etc. Through these imaging techniques, some cancerous symptoms of the body can be seen. But it is not feasible or available for everyone to detect cancer using all these techniques. so, there is a need of a affordable and easily available method to detect cancer at early stages.

How is this Deep learning project used to detect cancer??

Deep learning, a branch of machine learning, is an algorithm based on an artificial neural network to learn features from data. Deep learning proposes a method that enables computers to learn pattern features automatically. This property can be used to detect or prediction.

Model is trained on dataset(images) of cancer and non-cancer people. The model learns to identify patterns and features in the images that are indicative of cancerous or healthy conditions. Once trained, the model can analyse new tongue images and make predictions about the likelihood of cancer presence.

What are the future directions of your research in cancer detection using deep learning?

we aim to explore further to enhance the effectiveness and applicability of our deep learningbased approach to cancer detection. This includes investigating the integration of multi-modal imaging data, such as combining MRI, to improve diagnostic accuracy. Collaborations with clinicians and healthcare professionals will also be essential to validate the real-world effectiveness of our research and ensure its seamless integration into clinical practice.

Dataset link:-

https://www.kaggle.com/code/shivam17299/oral-cancer-lips-and-tongueimages-dataset/input https://www.kaggle.com/datasets/dssharadsharma/oral-health-data

The dataset consists of tongue classified into 2 groups cancer and non-cancer. The dataset is divided into 80/20 ratio of training and testing.

Methodology: -

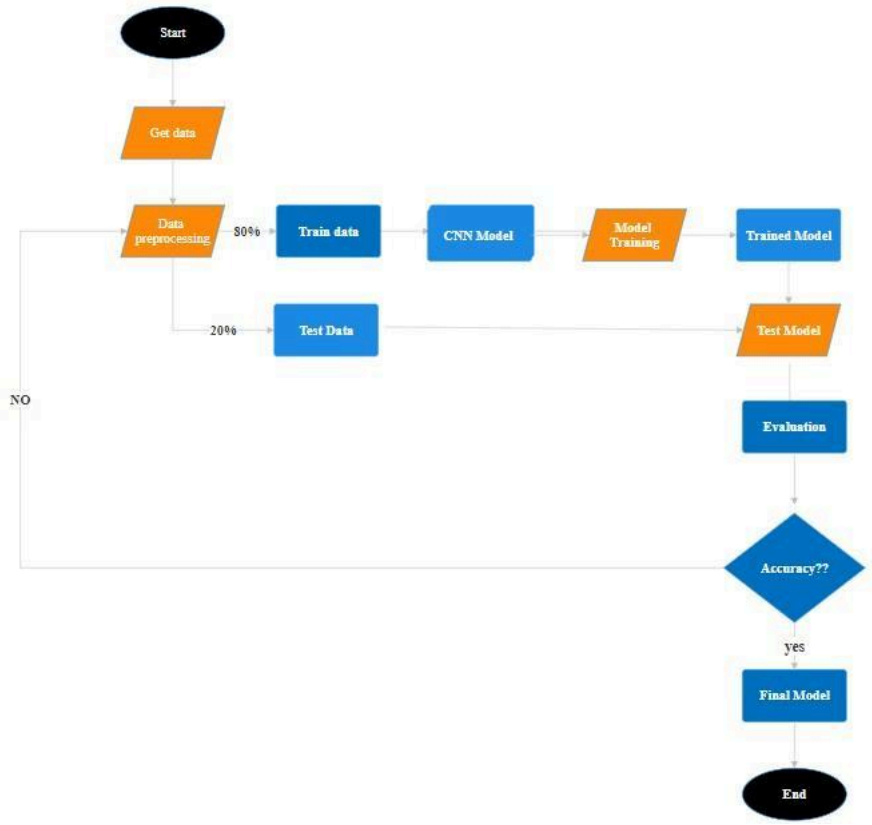

Fig: Flowchart

Get Data:

Obtaining medical imaging datasets containing annotated images of cancerous and noncancerous patients. Data has been collected from Kaggle whose link is mentioned.

Acquiring datasets with a sufficient number of samples is crucial to train a model effectively and ensure its generalization to unseen data. Total images in dataset is nearly 800 which further requires data preprocessing and needed to be divided into train and tested dataset.

Dataset link:-

https://www.kaggle.com/code/shivam17299/oral-cancer-lips-and-tongueimages-dataset/input

https://www.kaggle.com/datasets/dssharadsharma/oral-health-data

Data Preprocessing:

Data preprocessing plays a crucial role in preparing medical imaging data for model training. Perform preprocessing steps such as image normalization, resizing. Data set has been divided into two sets Train data (which contain 80%) and Test data (which contain rest 20% of data). Train data and test data are kept in two separate folders.

Train folders contain the two folder of cancer and non-cancer images and this similarly goes for test folder.

Model Selection:

Choosing the appropriate deep learning architecture is a critical decision that significantly impacts the performance of the cancer detection system. Convolutional Neural Network (CNN), based on the characteristics of the dataset and the specific requirements of the project. Various neural network like vgg can also be used for model selection . vgg is neural model which already trained on various images but here basic CNN model is used.

Model Training:

Model is trained based on the selected training data. Model training involves feeding the pre-processed data into the selected deep learning architecture. The images of train dataset folder is feed into the model. The model is trained on cancer and non-cancer images.

Test Model:

The trained model is tested based up on the test data set. Model testing involves Testing trained model on the basis of test dataset. The test data set will decide that whether model has good accuracy or not .

Model Evaluation:

Once the model is trained, its performance needs to be evaluated to assess its effectiveness in cancer detection. This evaluation involves testing the trained model on a separate test dataset, which contains unseen samples not used during training or validation. Evaluate the trained model's performance using the metrics such as accuracy, precision, recall, and F1-score.

Accuracy:

After the evaluation if the model is having good accuracy/desired accuracy then the model can be said as final trained model. If the model accuracy is less than the desired one, then process needs to be redone from data processing step. Once the model has demonstrated satisfactory performance during testing, it can be deployed as a computeraided diagnosis system for clinical use.

Output:

```
import numpy as np
from tensorflow.keras.preprocessing.image import load_img, img_to_array
from tensorflow.keras.models import load_model
from keras.applications.mobilenet import preprocess_input

class_labels = {1: 'NON_CANCER', 0: 'CANCER'}

def predict_image_class(image_path, model):
    img = load_img(image_path, target_size=(224, 224))
    img_array = img_to_array(img)
    img_array = preprocess_input(img_array)
    img_array = np.expand_dims(img_array, axis=0)
    predictions = model.predict(img_array)
    pred_class_index = np.argmax(predictions, axis=1)[0]
    pred_class_label = class_labels[pred_class_index]
    return pred_class_label, predictions[0]

image_paths = [
    "/content/drive/MyDrive/ML_project/Train/CANCER/496.jpeg",
    "/content/drive/MyDrive/ML_project/Train/CANCER/494.jpeg",
    "/content/drive/MyDrive/ML_project/Train/NON _CANCER/012.jpeg",
    "/content/drive/MyDrive/ML_project/Train/NON _CANCER/029.jpeg",
    "/content/drive/MyDrive/ML_project/IMG_20240422_235021.jpg",
    "/content/drive/MyDrive/ML_project/IMG_20240422_235311.jpg"
]

model = load_model("/content/best_model.h5")

for path in image_paths:
    pred_class_label, pred_probabilities = predict_image_class(path,
model)
    print("Image:", path)
    print("Predicted class:", pred_class_label)
    print("Predicted probabilities:", 1-pred_probabilities)
    print()
```

```
1/1 [==============================] - 1s 572ms/step
Image: /content/drive/MyDrive/ML_project/Train/CANCER/496.jpeg
Predicted class: CANCER
Predicted probabilities: [1.]

1/1 [==============================] - 0s 58ms/step
Image: /content/drive/MyDrive/ML_project/Train/CANCER/494.jpeg
Predicted class: CANCER
Predicted probabilities: [1.]

1/1 [==============================] - 0s 61ms/step
Image: /content/drive/MyDrive/ML_project/Train/NON _CANCER/012.jpeg
Predicted class: CANCER
Predicted probabilities: [0.]

1/1 [==============================] - 0s 62ms/step
Image: /content/drive/MyDrive/ML_project/Train/NON _CANCER/029.jpeg
Predicted class: CANCER
Predicted probabilities: [0.]

1/1 [==============================] - 0s 57ms/step
Image: /content/drive/MyDrive/ML_project/IMG_20240422_235021.jpg
Predicted class: CANCER
Predicted probabilities: [2.104044e-05]

1/1 [==============================] - 0s 55ms/step
Image: /content/drive/MyDrive/ML_project/IMG_20240422_235311.jpg
Predicted class: CANCER
```

Cancer and non-cancer images are given to model at the end for testing purpose.
Model predict as
Cancer = 0
Non-cancer =1
And 0 to 1 in between depending on prediction